

Sequence Homology Determination

(The example provided is aimed at finding a yeast ORF if given a rat accession number. This can be easily modified for other species combinations.)

Given the rat accession number: AA858888

- 1) Go into UniGene at <http://www.ncbi.nlm.nih.gov/UniGene>
- 2) Choose the rat species.
- 3) Enter the accession number and press Go.

You should get something like this:

The screenshot displays the NCBI UniGene web interface. At the top, the NCBI logo is on the left, and the UniGene title is on the right. Below the title is a navigation bar with links to PubMed, Entrez, BLAST, OMIM, Taxonomy, and Structure. A search bar contains the text 'Rat' and a 'Go' button. On the left side, there is a vertical menu with links to NCBI, UniGene, Home Page, Frequently Asked Questions, Query Tips, DDD-Library Digital Differential Display, Download UniGene, UniGene Rattus norvegicus, Home Page, Release Statistics, Library Report, Library Browser, DDD-Library Digital Differential Display, UniGene Organisms, and Homo sapiens. The main content area shows the UniGene Cluster Rn.2458 for Rattus norvegicus. It includes the title 'Rattus norvegicus mRNA for class I beta-tubulin, complete cds', a 'SEE ALSO' section with a link to HomoloGene: Rn.2458, and a table of 'SELECTED MODEL ORGANISM PROTEIN SIMILARITIES'. The table lists various organisms and their protein sequences with their percent identity and length. Below the table is an 'EXPRESSION INFORMATION' section with a note about representation and cDNA sources. At the bottom, the 'mRNA/GENE SEQUENCE (1)' section shows the accession number AB011679 and the title 'Rattus norvegicus mRNA for class I beta-tubulin, complete cds'. A black arrow points from the 'P/A' button to the accession number.

NCBI

UniGene

PubMed Entrez BLAST OMIM Taxonomy Structure

Search Rat Go

NCBI

UniGene

Home Page

Frequently Asked Questions

Query Tips

DDD-Library Digital Differential Display

Download UniGene

UniGene

Rattus norvegicus

Home Page

Release Statistics

Library Report

Library Browser

DDD-Library Digital Differential Display

UniGene Organisms

Homo sapiens

UniGene Cluster Rn.2458 *Rattus norvegicus*

Rattus norvegicus mRNA for class I beta-tubulin, complete cds

SEE ALSO

HomoloGene: [Rn.2458](#)

SELECTED MODEL ORGANISM PROTEIN SIMILARITIES

organism, protein and percent identity and length of aligned region

| | | |
|-----------------------|---|--|
| <i>H.sapiens</i> | : pir-I38369 - I38369 beta-tubulin | 98 % / 443 aa (see ProtEST) |
| <i>M.musculus</i> | : pir-E25437 - TUBULIN BETA-5 CHAIN | 100 % / 443 aa (see ProtEST) |
| <i>R.norvegicus</i> | : pir-A25113 - A25113 tubulin beta chain 15 - rat | 95 % / 442 aa (see ProtEST) |
| <i>A.thaliana</i> | : pir-JQ1587 - TUBULIN BETA-2/BETA-3 CHAIN | 85 % / 442 aa (see ProtEST) |
| <i>C.elegans</i> | : pir-T18683 - T18683 hypothetical protein B0272.1 - Caenorhabditis elegans | 93 % / 441 aa (see ProtEST) |
| <i>D.melanogaster</i> | : sp-P08840 - TBE2 DROME TUBULIN BETA-2 CHAIN | 94 % / 440 aa (see ProtEST) |
| <i>S.cerevisiae</i> | : pir-UBBYB - UBBYB tubulin beta chain - yeast (Saccharomyces cerevisiae) | 75 % / 440 aa (see ProtEST) |

EXPRESSION INFORMATION

Note: Highly represented (1.2 pct) in library [S5 UI-R-AA0](#)

cDNA sources: embryo, pluripotent cell line, other, whole embryo

mRNA/GENE SEQUENCE (1)

[AB011679](#) Rattus norvegicus mRNA for class I beta-tubulin, complete cds

P/A

4) Click under the mRNA/gene sequence to select clone. By clicking on the gene to get the sequence, you should get something like this:

NCBI **UniGene**

PubMed Entrez BLAST OMIM Taxonomy Structure

Search

NCBI
UniGene
Home Page
Frequently Asked Questions
Query Tips
DDD-Library Digital Differential Display

gnl|UG|Rn#S92767 2346 bases *Rattus norvegicus* UniGene [Rn.2458](#)

Rattus norvegicus mRNA for class I beta-tubulin, complete cds

FEATURES

| | | |
|------------------|------------------|---|
| db_source | 1 ... 2346 | GenBank entry: AB011679 |
| CDS | (+) 156 ... 1490 | Product: class I beta-tubulin See protein sequence PID:g3551054 |
| prot_sim | (+) 157 ... 1488 | <i>Homo sapiens</i> I38369 beta-tubulin (98% identity in aligned region, p=0.0) See protein sequence pir:I38369 Also See: ProtEST |

5) Go ahead and click on the GenBank entry to get the DNA sequence.

NCBI **Nucleotide**

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy

Search for

Limits Preview/Index History Clipboard

☐ 1: AB011679. *Rattus norvegicus*...[gi:3551053]

[Related Sequences](#), [Protein](#), [Pub](#)

```
>gi|3551053|dbj|AB011679.1|AB011679 Rattus norvegicus mRNA for class I beta-tubulin, c
CTCGCATCTCTGTCAGCAGCCTCCGGGGCTACGTTCCATCCTATCAGTCTGAGACCAGCCAGAAAGAAA
AAAACCAAAAATCCTTAATTATCTTCTTTTGCTCGGTACCTACATTGGGACCATCAAAAAAATTATTA
CAGTAAACCGTAGCCATGAGGGAAATCGTGACATCCAGGCCGACAGTGTGGCAACCAGATCGGTGCTA
AGTTCTGGGAGGTGATAGCGATGAACACGGCATCGACCCACCGGCACCTACACGGAGACAGCGACTT
GCAGCTGGACCGAATCTCTGTGTAACAATGAAGCTACAGGTGGCAAGTATGTCCCTCGAGCTATCTTA
GTGGATCTAGAACCCGGGACTATGGACTCCGTTCCGCTCAGGTCTTTTGGCCAGATCTTCAGACCGGACA
ACTTTGTTTTTGGTCACTCTGGGCAAGGCAACAAGTGGCTAAGGGTCACTACAGAGGGAGCTGAGCT
GGTGAATCTCTGCTTGGATGTGGTGCGGAGGAGGCGGAGAGCTGTGACTGCCTGCAAGGCTTTGAGCTG
ACCCAATCGCTGGGTGGAGGCAAGGCTCTGGCATGGGCAACCTGCTCATCAGCAAGATTTCGAGAAGAT
ACCCGACCGCATCATGAACACCTTCAGCGTGGTGCCTCAACCAAGTGTCTGACACCGTGGTTGAGCC
CTACAACGCCACCTGTCCGTCCACCAAGTTGGTAGAACAACAGATGAGACCTACTGCAATTGACAAACGAG
GCTCTCTACGACATCTGCTTCCGTACCCCTTAAGCTCACCACGCCAACCTACGGAGACCTGAACCATCTCG
```

6) On the next page, change the viewing options from GenBank to FASTA format and hit “display”. You should get the desired sequence that you want to blast.

7) In the case of blasting against yeast, go to this website: <http://genome-www2.stanford.edu/cgi-bin/SGD/nph-blast2sgd> and paste your sequence in the box shown below:

The screenshot shows the top of the 'S. cerevisiae WU-BLAST2 Search' web page. It features a red 'SGD' logo, a 'Help' button, and a link to 'Send a Message to SGD'. Below this is a 'Query Comment' text box. A note states: 'If sequence is less than 30 letters you need to set the S value to less than 100 or you can miss matches.' There is a section for 'Upload Local TEXT File' with a 'Browse...' button. A large text area is labeled 'Paste sequence here'. Below that is a dropdown menu for 'Choose the Appropriate BLAST Program:' with 'BLASTN : nucleotide query to nucleotide db' selected. Another dropdown menu for 'Choose a Sequence Database. All datasets contain only S. cerevisiae sequences.' has 'Complete genomic sequence, chromosomes (DNA)' selected. At the bottom are 'Run BLAST' and 'Reset' buttons.

S. cerevisiae WU-BLAST2 Search

Your comments and suggestions are appreciated : [Send a Message to SGD](#)

Query Comment (optional, will be added to output for your use):

If sequence is less than 30 letters you need to set the S value to less than 100 or you can miss matches.

Upload Local TEXT File: FASTA, GCG, and RAW sequence formats are okay
WORD Documents do not work unless saved as TEXT.

Browse...

Type or Paste a Query Sequence : (No Comments, Numbers Okay)

Paste sequence here

Choose the Appropriate BLAST Program:
BLASTN : nucleotide query to nucleotide db

Choose a Sequence Database. All datasets contain only *S. cerevisiae* sequences.
Complete genomic sequence, chromosomes (DNA)

Run BLAST or Reset

8) Under the “choose the appropriate BLAST program” go ahead and choose the appropriate BLAST Program:

This is a close-up of the 'Choose the Appropriate BLAST Program:' dropdown menu. The selected option is 'BLASTN : nucleotide query to nucleotide db'.

Choose the Appropriate BLAST Program:

BLASTN : nucleotide query to nucleotide db

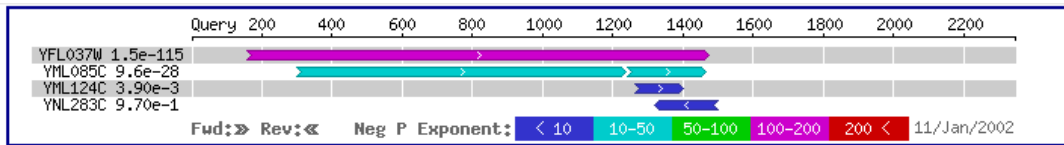
9) Under Sequence database, choose the appropriate database:

This is a close-up of the 'Choose a Sequence Database. All datasets contain only S. cerevisiae sequences.' dropdown menu. The selected option is 'ORF DNA, Coding Sequences of defined ORFs (DNA)'.

Choose a Sequence Database. All datasets contain only *S. cerevisiae* sequences.

ORF DNA, Coding Sequences of defined ORFs (DNA)

10) Press Run Blast. You should get something like this, giving you yeast ORF identifiers for top hits with some vital statistics:



| | | | High | Smallest | |
|---|---|------|--------------------------|-------------|---|
| | | | Score | Sum | |
| Sequences producing High-scoring Segment Pairs: | | | | Probability | N |
| | | | | P(N) | |
| YFL037W | TUB2, Chr VI from 56335-57708 | 2640 | 1.5e-115 | 1 | |
| YML085C | TUB1, Chr XIII from 97941-97965,98082-99400,... | 656 | 9.6e-28 | 2 | |
| YML124C | TUB3, Chr XIII from 22049-22073,22372-23684,... | 249 | 0.0039 | 1 | |
| YNL283C | WSC2, Chr XIV from 105182-106693, reverse co... | 186 | 0.97 | 1 | |